Police agencies on Facebook overreport on Black suspects

Ben Grunwald
Julian Nyarko
John Rappaport

Follow this and additional works at: https://chicagounbound.uchicago.edu/journal_articles

Part of the Law Commons
A large and growing share of the American public turns to Facebook for news. On this platform, reports about crime increasingly come directly from law enforcement agencies, raising questions about content curation. We gathered all posts from almost 14,000 Facebook pages maintained by US law enforcement agencies, focusing on reporting about crime and race. We found that Facebook users are exposed to posts that overrepresent Black suspects by 25 percentage points relative to local arrest rates. This overexposure occurs across crime types and geographic regions and increases with the proportion of both Republican voters and non-Black residents. Widespread exposure to overreporting risks reinforcing racial stereotypes about crime and exacerbating punitive preferences among the polity more generally.

Most Americans learn about crime through the media rather than personal experience (1). News sources, in particular, powerfully shape public perceptions about crime and, by extension, crime policy (2–4). Today, a majority of American adults get news from social media platforms, Facebook most common among them (5, 6). In that forum, law enforcement agencies communicate directly with the public about crime. Indeed, Facebook’s own user guide encourages agencies to “[l]et the community know when to look for suspects and when they have been apprehended” (7). Parsing the messages agencies send on social media illuminates how the arm of the state principally responsible for enforcing laws sends on social media illuminates how the arm of the state principally responsible for controlling crime also shapes public views about its character and incidence.

With this goal in mind, we gather all posts from almost 14,000 Facebook pages maintained by US law enforcement agencies to study how agencies communicate with the public about crime. Given the centrality of race to public policy and discourse about policing and crime, we identify nearly 100,000 posts that report on the race of individuals suspected of or arrested for crimes. We estimate these posts received 197 million exposures to followers who, in turn, shared them more than 11 million times with their own Facebook communities (see SI Appendix, section S1 for details). By comparing the relative frequency of agency posts about Black suspects with local and aggregate arrest statistics, we examine whether law enforcement communications on Facebook expose users to an accurate representation of the interaction between crime and race or whether instead they reproduce, and threaten to entrench, stereotypes about “Black criminality” (8).

Prior research has examined whether and how traditional media overreport on crimes involving Black suspects. Early work on local newspapers and TV news in a small number of locales found evidence of overreporting (9–12). It also showed that exposure to the “racial crime script” promotes negative attitudes toward Black people (1)—especially among white residents of racially homogeneous communities (13)—as well as support for Republican political candidates (14). Other research, mostly more recent studies and studies based on small samples of nationally syndicated television news programs, has found little evidence of overreporting (15–18). One recent paper finds that homicides involving racial minorities—as either perpetrator, victim, or both—receive substantially more local news coverage than white-on-white homicides (19).

The ashen of social media as a source of crime news requires a ground-up rethinking of this issue. Whereas traditional media can constrain and filter how law enforcement communicates with the public, social media has no external gatekeepers (6, 20, 21). Instead, law enforcement itself decides when and how to report on crime (22). The bureaucratic process for selecting content varies across agencies—some agencies empower a single individual, for example, while others require preapproval by a central group (23). A related literature on media economics suggests that content filtering can be shaped by both supply-side and demand-side factors, such as the ideological preferences of the content producers and consumers, respectively (24, 25).

The effects of news consumption through social and traditional media also may differ. A substantial body of research on traditional media, for example, documents a

**Significance**

By analyzing all Facebook posts from pages maintained by US law enforcement agencies, we show that Facebook users are exposed to posts about serious crimes that significantly overrepresent Black suspects relative to local arrest rates. Our results point to one mechanism by which the state itself may reinforce racial stereotypes about crime. Prior research shows that these stereotypes, in turn, may heighten demand for the state’s penal services.
positive correlation between consumption of crime news and fear of crime (for a review of the literature, see ref. 26). 

One recent study found that this relationship is even stronger on social media, perhaps because of the more active nature of reader engagement on these platforms (26). For similar reasons, consumption of news about Black criminal suspects on social media may exacerbate racial stereotypes particularly acutely. The “incipient and fragmentary” body of research about law enforcement social media activity does not address these questions (27).

Data

We constructed our Facebook dataset using CrowdTangle (28), a website that tracks interactions on public content from Facebook pages and groups (see, e.g., ref. 29). In addition to the posts themselves, we extracted post-level metadata such as the number and types of user interactions and the number of page followers. Using Google Maps, we associated posts with the geolocation of their originating agencies. We then used several algorithms to identify posts that describe both a crime and the geolocation of their originating agencies. We then used several algorithms to identify posts that describe both a crime and the geolocation of their originating agencies.

Finally, we matched agency Facebook pages, and by extension their posts, to agencies in the Federal Bureau of Investigation (FBI)’s Uniform Crime Reporting (UCR) program, the most commonly used dataset in research about crime. We merged in additional data from the UCR, including annual arrest data by race (30). From the universe of all 100,000 race–crime posts, we analyze the nearly 70,000 matched posts about serious, part I offenses committed—because agencies exercise discretion in whom to arrest and how to classify their offenses—arrests are the best available proxy. Note that if Black suspects are more likely than white suspects to be arrested for similar crimes, this renders the estimates in our analysis conservative. Nevertheless, to address concerns about possible bias, we show that our results are robust to using offense-level data from the FBI’s National Incident-Based Reporting System (NIBRS). Although this dataset records offenses directly, it covers fewer agencies than the UCR and is missing race information for a large number of crimes and suspects. For more details on our dataset and the algorithms we used to construct it, see Materials and Methods.

Results

Agency Overreporting. We first examine whether agencies overreported on Black suspects in race–crime posts relative to local arrest statistics. We define an agency’s overreporting score as the difference between the proportion of the agency’s arrestees who are Black and the proportion of Black suspects in the agency’s race–crime posts (for a formal definition, see Materials and Methods). To estimate these agency-level scores, for each agency we collapse together all posts between 2010 and 2019. Taking the average of our agency-level estimates, Black suspects were described in 32% of race–crime posts but represented just 20% of arrestees (Fig. 2). Overreporting is both substantively and statistically significant (SI Appendix, Fig. S1, Top Left) for each individual crime category except auto theft, suggesting prevalent disparities in agencies’ Facebook posting activity even for the most reliably reported offenses like homicide and robbery.

Without more, however, establishing that the average agency overreports on crimes involving Black suspects tells us relatively little about the overall impact of agencies’ posting activities on the American public. For one, users can learn about and pay attention to reported crimes that occur outside of their own agency’s jurisdiction. In addition, the agencies exhibiting the most significant disparities might rarely post on Facebook, or
their pages may have few followers. We turn next to examining patterns in exposure to overreporting, a construct that accounts for these considerations.

**Overexposure to Posts about Black Suspects.**

**Defining overexposure.** To estimate exposure to posts about Black suspects, we assume that while users sometimes encounter crime reports from outside their own agency’s jurisdiction, they are most likely to learn about nearby incidents. Two basic observations support this assumption. First, the intensity of Facebook friendship networks strongly declines in geographic distance; on average, over 60% of Facebook friendship links are to people living within 100 miles (32). Accordingly, when a user’s Facebook friends share or mention race–crime posts from agencies they follow, those posts are more likely to come from nearby agencies than distant ones. Second, physical proximity heightens the impact of crime on fear of future crime; specifically, crime in the immediate area seems most impactful, but crime in surrounding regions matters too (33). For this reason, many users may follow or learn about not only their local agency’s page but also the pages of other, nearby agencies; they are less likely, however, to do so for more distant agencies.

We thus develop an agency-level measure of distance-weighted overexposure to crime posts about Black suspects. Because Facebook users can be exposed to posts from agencies other than their own, we calculate overexposure for all agencies that report arrest information, including those that do not have a Facebook page. We begin by computing, for each agency, the proportion of arrestees who were Black using all arrests occurring between the years 2010 and 2019. Next, we calculate for each agency the distance-weighted proportion of posts that described a Black suspect. For this, we use all posts by agencies within a 300-mile radius. Our weights are distance decay spatial weights for each agency, meaning that posts from a jurisdiction’s own agency receive a weight of 1, while other posts are exponentially discounted based on their distance to that agency (see, e.g., ref. 34; for a formal definition of the decay function, see Materials and Methods). In our primary specification, for example, the decay function assigns weights from agencies 50 or 100 miles away weights of 0.7 and 0.3, respectively (SI Appendix, Fig. S2).

To arrive at our agency-level measure of overexposure, we then subtract the proportion of arrests involving Black suspects from the weighted proportion of posts about Black suspects (for a formal definition, see Materials and Methods). We use arrests from the local agency but not surrounding agencies because our aim is to assess how well the crime news to which users are exposed—which includes news posted by surrounding agencies—matches the real-life crime patterns in their local communities. Finally, we average the agency-level overexposure scores, giving equal weight to each agency.

Our measure of overexposure is necessarily imperfect. It assumes that the probability of exposure to a post at a given distance is constant across the United States even though it is possible, for example, that a user in a sparsely populated area is more likely to observe a post from an agency 50 miles away than a user in a densely populated region. We also cannot directly observe which users viewed which posts. As we report below, however, our results are robust to various alternative specifications, including using alternative distance decay parameters and different ways of weighting agencies’ posts in calculating aggregate overexposure. We also show robustness to post-weighting schemes that do not rely on distance-weighting but instead on the number of followers and interactions associated with each post.

**Estimating overexposure.** Fig. 3 compares exposure to posts about Black suspects with the proportion of Black arrestees, averaged across all jurisdictions. In SI Appendix, we separately plot the difference—our measure of overexposure—with a 95% confidence interval (SI Appendix, Fig. S3, Left) obtained from an equivalent regression analysis (SI Appendix, section S2). When aggregating all crimes together, we estimate that users were exposed to race–crime posts that overrepresented Black suspects by 25 percentage points (95% CI 24 to 25) relative to local arrest rates. Given that the average proportion of Black arrestees is 18 percentage points, this constitutes relative overrepresentation of 138%.

One potential explanation for this aggregate overexposure is that within any given crime type, agencies are more likely to post about Black suspects than suspects of other racial groups. Because the racial composition of arrestees varies substantially across crime types, however, it is also possible that our estimate of aggregate overexposure is driven by agencies reporting more frequently on certain crime types, such as those thought particularly salient to residents’ perceived safety. If Black people represent a large proportion of arrestees for these crime types (relative to other crime types), this could explain the overexposure we observe. For example, on average, 34% of arrests for robbery in our sample are Black, whereas Black suspects make up only 20% of individuals arrested for aggravated assault. If we observe overexposure in the aggregate across these two crimes, we cannot tell whether agencies are reporting more frequently about Black robbery suspects than white robbery suspects or, alternatively, whether they are reporting more frequently on robberies in general (as compared to aggravated assaults). To determine whether the latter possibility could fully explain our results, we examine overexposure within individual crime types.

As shown in Fig. 3, overexposure was most pronounced for violent offenses—where racial stereotypes about Black criminality are perhaps most damaging. Facebook users were exposed to violent crime posts in which Black suspects appeared 26 percentage points (95% CI 25 to 27) more often than Black arrestees appeared in local arrest data—a relative difference of...
121% over the average Black arrest rate. Notably, overexposure was both statistically and substantively significant among all violent crimes, ranging from 18 to 21 percentage points. Overexposure was also prevalent among property crimes at an average of 16 percentage points (95% CI 15 to 17).

That we find substantial overexposure within each crime type rules out the possibility that aggregate overexposure was driven entirely by agencies reporting more frequently on crimes for which Black people represented a larger fraction of arrestees. Moreover, in SI Appendix, Table S1, we provide further evidence that overexposure to Black suspects within the various crime categories drove a large fraction of the aggregate overexposure.

We demonstrate the robustness of our results from Fig. 3 in several ways. First, the results are substantively identical when we use alternative distance decay parameters in calculating overexposure for each agency (SI Appendix, Fig. S3, Middle Right).

Second, the results are also robust to an aggregation process that does not assign equal weight to each agency’s estimated overexposure but instead weights agency overexposure by the size of the population each agency serves (SI Appendix, Fig. S4). Because many law enforcement agencies serve overlapping jurisdictions, measuring the population served by each agency is not straightforward. We use the UCR’s agency-level population estimates as a plausible, albeit imperfect, measure (for more details on how the UCR estimates populations for overlapping municipal and county agencies, see SI Appendix, section S4). When we aggregate across agencies this way, average overexposure is somewhat lower than in Fig. 3 but still substantial, corresponding to relative overrepresentation of Black suspects of 50%.

Third, the results are robust to the use of different post-weighting schemes altogether, which do not rely on distance weighting at all and include only agencies with Facebook pages. For example, because posts from pages with more followers, as well as posts that receive more engagement (interactions or shares), may reach a higher number of users, we estimate overexposure using only the subset of race–crime posts with the least investigative value—those that report on the arrest of a Black suspect (SI Appendix, Fig. S6).

The Geography and Demography of Overexposure. We next examine the prevalence of overexposure across socially salient geographic and demographic dimensions. Fig. 4 depicts the national distribution of distance-weighted local overexposure, interpolating values in areas not covered by our dataset (see SI Appendix, Fig. S7 for a map without interpolation). Most of the country was exposed to overreporting on Black suspects, with much of the Midwest and some of the South and mid-Atlantic regions experiencing it most intensely. Half of jurisdictions experienced overreporting of 26 percentage points or greater (SI Appendix, Fig. S8). The only areas where users were not consistently exposed to overreporting were Hawaii and the Black Belt in the South, a region that relied heavily on slave labor until the Civil War and that retains many Black population centers today (35). Agencies in Florida and Illinois do not systematically report arrest data to the UCR. Our predictions in these states are thus based on few observations (two and seven, respectively).

Consistent with past work finding that Republicans and Democrats are increasingly exposed to different news (36), we next examine whether overexposure to race–crime posts about Black suspects varies with political partisanship. To measure political preferences at the county level, we average the county-level Republican vote share over all presidential elections during our period of observation. Fig. 5, Top, shows that overexposure increased, on average, with the share of Republican voters in the county—from roughly 0 overexposure where Republican candidates received 25% of the vote to around 25 percentage points where they won 75%.

Republican vote share correlates strongly with the racial distribution of the population (38). The observed relationship between political preferences and overexposure might therefore be explained by the presence of larger Black populations in Democratic counties. We explore this relationship, plotting local overexposure by the proportion of Black residents in the county. As Fig. 5, Bottom, shows, overexposure was indeed higher in counties with fewer Black residents. This was true in both majority-Republican and majority-Democrat counties. At the same time, even when Black population shares were similar, overexposure was substantially higher in majority-Republican counties. Where Black residents made up 5 to 25% of the population, for example, jurisdictions in majority-Republican counties experienced roughly 15 to 20 percentage points more overexposure than jurisdictions in majority-Democrat counties.

Discussion

Our analysis determined that Facebook users were exposed to posts that overrepresented Black criminal suspects by about 25 percentage points relative to local arrest rates—26 percentage points for posts about violent crime and 16 for property crime. Our findings thus contrast with recent work uncovering little evidence of overreporting on Black criminal suspects and defendants in traditional media like TV and print news outlets (17, 18). Overexposure on Facebook was present across the country, the only notable exceptions lying in Hawaii and the Black Belt, a Southern region with large Black population centers. At the same time, the intensity of overexposure correlated strongly with the average Republican vote share in presidential elections.
In the context of racialized media coverage, the literature has proposed several possible explanations for the existence of disparities (24). As these explanations focus principally on the reporting practices of content producers—rather than on the content to which consumers are exposed—we probe them using our measure of overreporting rather than overexposure. First, disparities in reporting may be supply-driven, reflecting the preferences or incentives of the content producer. Similar to Fig. 5, Top, we find that overreporting correlates with county political preferences (SI Appendix, Fig. S9, Top), evidence consistent with this theory, as more politically conservative counties may employ more conservative police officers, who, in turn, may be more racially conservative (39–41). At the same time, we find no relationship between overreporting and the share of Black officers at each agency (SI Appendix, Fig. S10). We also find little evidence that pressures external to the

Fig. 5. Overexposure by county-level characteristics. (Top) Scatterplot of county-level overexposure by the average of county-level Republican vote share in the 2012 and 2016 presidential elections. (Bottom) Scatterplot of county-level overexposure by the county-level proportion of Black residents, based on Census counts. We use Census data from 2010, the most recent available figures that do not rely on intercensal estimates. Election data are from ref. 37. In both graphs, overexposure is winsorized to increase readability. Loess curves with a smoothing parameter of 0.75 indicate weighted average overexposure rates, separately for majority Republican- and Democrat-voting counties in Bottom. The gray-shaded areas indicate the 95% confidence intervals of the loess curves.
agency drive supply-side mechanisms. For example, while racial threat theory might predict that overreporting would rise with Black population share as officers perceive a growing community threat (42), we find little evidence of such a trend (SI Appendix, Fig. S9, Bottom). We also find no evidence that direct electoral pressures are contributing to the disparities, as there are no differences in overreporting between sheriffs’ offices, the leaders of which are typically elected, and police departments, where chiefs are usually appointed (SI Appendix, Fig. S11).

Second, disparities may be demand-driven, reflecting the preferences of content consumers. Notably, one recent study points to a demand-side mechanism to explain a racialized skew in local news coverage of homicides (19). Preference heterogeneity is well documented across both political and racial groups (25, 43), making this theory a plausible fit for the relationships just mentioned between overreporting and politics and race, respectively.

Third, in our particular setting, overreporting and, in turn, overexposure could potentially arise from agency emphasis—for either supply- or demand-driven reasons—on crimes for which Black people are disproportionately arrested. While evidence presented above is generally inconsistent with this explanation (Fig. 3 and SI Appendix, Table S1), it accounts only for differences in the type of crime committed. We cannot rule out that crime characteristics may differ in other, unobservable ways that could help explain the disparities.

Our analysis is descriptive in nature and thus is not designed to fully identify the causal mechanisms that led to the racial disparities in exposure. This means that at bottom, we cannot determine whether the observed disparities were driven by racial animus, implicit bias, or other, facially neutral factors. In the United States, this distinction carries legal significance. Racially animated actions by public actors constitute disparate treatment, a form of discrimination outlawed under the Equal Protection Clause of the Fourteenth Amendment. In contrast, actions that are not racially animated but nonetheless have adverse consequences for a protected group can constitute disparate impact, a form of discriminatory conduct deemed illegal by federal and state laws only in isolated contexts (such as employment, housing, or credit decisions).

A legally significant distinction, however, is not necessarily normatively meaningful or policy-relevant. Even if the observed disparities resulted from facially neutral actions, they may impose substantial social costs. As one study summarizes the evidence—“each instance of exposure to a Black criminal activates the stereotype of Black criminality, strengthening the cognitive association between Blackness and crime” (ref. 44, p. 107; see also ref. 1). The impact of crime news on racial attitudes seems particularly pronounced for white people in white homogeneous communities (13)—precisely where exposure to overreporting is highest. The stereotype of Black criminality, in turn, can have harmful effects, animating biased treatment of Black people in domains ranging from criminal justice decision-making (45, 46) to employment (47) to the decision to use gun violence (48). Previous findings suggest that the association of crime with Blackness can also strengthen public support for punitive criminal justice policies (1, 13, 49, 50), support that can influence policy on the ground (51, 52) through the decision-making of legislators (53), prosecutors (54), and elected judges (55).

The practice of law enforcement agencies posting news about local crime may also produce some social benefits. Among other things, Facebook posts may, on occasion, allow agencies to enlist the community’s help to apprehend potentially harmful suspects (27). Our findings, however, remain even when we analyze only race–crime posts that lack this potential investigative value—those that announce a suspect’s arrest (SI Appendix, Fig. S6). Any benefits, moreover, must be weighed against the real costs of the practice, both for the Black community and for society at large. Agencies might therefore opt to report—and especially to report on suspect race—only where the expected benefits of doing so are demonstrably significant, as might be the case for serious crimes where a suspect remains at large (9).

For more minor offenses, or when a suspect is in hand, it is difficult to see how the benefits could outweigh the costs.

Materials and Methods

Our analysis combined two core datasets. First, we created a novel dataset on crime reports published on Facebook by law enforcement agencies across the US. This dataset documents how law enforcement informs the general public about local crime. Second, we incorporated data from the FBI’s UCR program to assess how those agency communications compared to the reality of local arrest statistics.

Identifying Facebook Pages. We compiled a list of US law enforcement agencies’ official Facebook pages in two stages. First, we used CrowdTangle, a website run by Facebook that tracks interactions on public content from Facebook pages, to search the titles of all public pages for common keywords in agency names, such as “police” and “sheriff.” Through this process, we identified 12,071 law enforcement agency pages. We excluded pages for police unions, fraternal organizations, auxiliary police departments, volunteer search-and-rescue teams, state-level agencies, and individual law enforcement officers.

In the second stage, we matched these Facebook pages with their respective UCR agencies primarily based on page and agency names and state locations (for the details of our matching procedures, see SI Appendix, section S3). In total, we matched 11,037 Facebook pages to 10,395 UCR agencies, 9,738 of which reported arrest information to the UCR during the study period.

We then conducted a manual search for additional Facebook pages associated with any unmatched agencies that reported arrest information to the UCR during the study period and have a population value greater than 0 recorded in the UCR. We found 1,919 additional pages through this process, which brought our final list to 13,990 pages, of which 12,971 were matched to 12,324 agencies. A small number of agencies—especially large agencies—have more than one Facebook page.

Processing Facebook Posts. Next, we used CrowdTangle to extract the text of all 11,058,289 public Facebook posts published before 1 January 2020 from all pages on our list, including both pages matched to UCR agencies and those not matched.

To identify posts containing one or more descriptions of an individual’s race, we created a list of keywords and keyword stems commonly used to identify race: alnca, arab, asian, biracial, black, brown, caucasian, hispanic, latin, mexic, skinned, white. We then adopted an algorithm developed by ref. 56 to identify race descriptions. We applied this algorithm to all posts that included at least one of our keywords to determine whether the posts described a person’s race (rather than, say, the color of a car) and identified 107,889 such posts. We confirmed through manual review that this approach correctly identified race descriptions with an accuracy of 98% (F1 score of 0.90).

To identify the subset of posts with race descriptions that also contained one or more descriptions of a crime—what we call crime–race posts—we hand-labeled a random sample of 990 posts. We recorded whether each post described a crime and, if so, whether it described one or more of the seven UCR part I offenses, for which arrest data are most reliably collected: murder (including non-negligent manslaughter), rape, robbery, aggravated assault, burglary, automobile theft, and theft. We also recorded whether each post reported a suspect’s arrest.

Consistent with best practices in the field, we used three types of approaches to classify the labels. In the first, the text of each post was represented as a series of dummy (or “one-hot encoded”) vectors. These vectors were then fed into several machine learning classifiers (ADABOOSTING, Gradient Boosting, Bernoulli Naïve Bayes classifier, and Multinomial Naïve Bayes). In a second approach, we fine-tuned a pretrained language model (Bidirectional Encoder Representations from Transformers, or “BERT”) (57) to predict each
label. A third approach used keywords, which was necessary for the "rape" label because none of the posts in our hand-labeled dataset were explicit about the actions perpetrated by a suspect of a sexual assault. Because of the narrow offense definition in the UCR, in implementing our keyword approach, we opted to label a post as describing a rape only if it explicitly mentioned the word "rape" or variations of the phrase "sexual assault in the first degree."

To address concerns for overfitting, we used 10-fold cross validation. In effect, this means that we trained each classifier on our hand-labeled data 10 times, holding back a different set of 10% of our data each time, which we used to test the classifier’s performance. We then compared the average area under the receiver operating characteristic curve across all 10 folds for each classification approach and selected the one with the highest performance for each label (SI Appendix, Table S2).

Ultimately, we identified 98,448 total race-crime posts and matched 70,310 race-crime posts about part I crimes, which we use in our analysis (for a complete description of each of our sample restrictions for posts, see SI Appendix, Table S3). While the race-description algorithm we applied cannot distinguish between posts describing the race of a suspect and posts describing the race of another individual, such as a victim, the latter were exceedingly rare within our final sample. Rare as well were race-crime posts that described multiple criminal incidents involving different suspects. A small number of posts, roughly 4%, describe multiple individuals of different races. We count a post as describing a Black suspect if at least one suspect is described as Black.

The number of these posts in our analytical sample, as well as the number of unique pages publishing them, steadily increased during our sample period. The cumulative number of posts grew from 390 in 2010 to 70,310 in 2019 (SI Appendix, Fig. S12, Left). Similarly, the cumulative number of agency pages with at least one race-crime post rose from 141 to 5,925 during that period (SI Appendix, Fig. S12, Right). In 68% of all part I crime-crime posts, the most serious offense was a violent crime, while in 32%, it was a property crime. Roughly 47% described a Black suspect.

We also collected metadata for each race-crime post in our analysis, such as the number and type of interactions and the number of page followers at the time the post was published. CrowdTangle tracks follower counts for Facebook pages after they are added to the CrowdTangle platform. The median date at which tracking began for all agency pages in our dataset is 23 December 2014, resulting in missing follower counts for 39% of our posts. We imputed missing follower counts by training a global forecasting model using NEATs (58) on all agency-level time series. We relied on the darts package in Python, v0.9.0 (https://github.com/unit8co/darts) to train the model. Where possible, we also associated each post with the geolocation of the originating agency through Google Maps. We were able to geolocate 13,054 out of 13,062 agencies in our sample.

Finally, to allow us to compare crime reports on Facebook to actual arrest statistics, we merged data on the number of arrests agencies reported to the UCR from 2010 to 2019 (ref. 30; for more details, see SI Appendix, section S4). We focus on Black arrestees because arrest data for other nonwhite racial and ethnic groups are either too sparse or too unreliable for our analysis. Some law enforcement agencies do not report to the UCR. To test whether our results are driven by reporting agencies being unrepresentative of nonreporting agencies, we run a robustness test on municipal agencies serving populations of 10,000 people or more (SI Appendix, Fig. S13)—a subset of agencies in which reporting and nonreporting agencies appear to be similar in population size, a variable strongly correlated with crime (see ref. 59). Furthermore, some reporting agencies do not report in all years. Our results are robust to including only agencies that report at least one arrest in every month from 2010 to 2019 (SI Appendix, Fig. S14). To supplement our primary analysis, we also show robustness to using NIBRS data from ref. 60 (for more details, see SI Appendix, section S5).

Analytical Strategy. This section provides formal definitions for several of our principal measures. As explained in the Results section of the main text, we first define an agency’s overreporting score as the difference between the proportion of the agency’s arrestees who are Black and the proportion of Black suspects in the agency’s race-crime posts. Formally, an agency’s overreporting, $\gamma$, is defined as

$$\gamma = \frac{1}{N_p} \sum_p l_p(\text{black}) - \frac{1}{N_s} \sum_s l_s(\text{black})$$

where $p \in 1, 2, \ldots, N_p$ are the posts of the agency, $a \in 1, 2, \ldots, N_s$ are reported arrests, and $l_{p,a}(\text{black})$ equals 1 if the post/arrest is about a Black suspect and 0 otherwise.

We then define an agency-level measure of user overexposure to posts about Black suspects. This measure incorporates distance-weighted posts from all agencies within 300 miles. In our primary specification, the decay function is

$$w_d = e^{-0.00005d^2}$$

where $d$ is the distance measured in kilometers. Our results are robust to using alternative distance decay parameters, however (SI Appendix, Fig. S3).

Formally, then, an agency’s overexposure, $\theta$, is defined as

$$\theta = \frac{1}{\sum_{p \in S_d} w_p} \sum_{p \in S_d} l_p(\text{black}) - \frac{1}{N_s} \sum_s l_s(\text{black})$$

where $a \in 1, 2, \ldots, N_s$ are reported arrests and $l_{p,a}(\text{black})$ equals 1 if the post/arrest is about a Black suspect and 0 otherwise.

As an alternative to distance-weighting, we estimated overexposure using three different measures of user engagement for each agency: the mean of shares across all agency posts, the mean of followers at the time posts were published, and the mean of interactions (i.e., like, haha, love, wow, sad, and angry). We calculated each agency’s overreporting score, $\gamma$. We then multiplied these overreporting scores by one of the measures of engagement. Formally, an agency’s overexposure is $\delta_{i,s}$ defined as

$$\delta_{i,s} = w_s \gamma_{i,s}$$

with

$$w_s = \frac{1}{N_p} \sum_p y_{p,s}$$

where $s$ indicates the metric to weight on, $p \in 1, 2, \ldots, N_p$ are the agency’s posts, and $y_{p,s}$ represents the post-specific metric (e.g., number of interactions). Results are displayed in SI Appendix, Fig. S1.

Data, Materials, and Software Availability. As described in Materials and Methods, we downloaded Facebook posts from CrowdTangle at https://crowdtangle.com/ (28). Arrest (30) and NIBRS data (60) were downloaded from Jacob Kaplan’s files on OpenCPSR, available at https://www.opencpsr.org/opencpsr/project/102263/view and https://www.opencpsr.org/opencpsr/project/118281/view (version V5) at https://doi.org/10.7910/DVN/NKLYWK (61). Due to restrictions from CrowdTangle, we are unable to provide the raw text, number of interactions, number of shares, or the subscriber count for posts in our data. Users interested in this information may collect it directly from CrowdTangle through the post identifiers we have included in the replication file.

ACKNOWLEDGMENTS. We thank Alex Choulas-Wood, Sharad Goel, Jacob Goldin, Joshua Grossman, Colleen Honigsberg, Aziz Huq, Jonathan Masur, Sarath Sanga, Roshena Sommers, and Keniel Yao for helpful suggestions and DeRay Mckesson and Naomi Shiffman for data support. Dylan Baker, Michael Berg, Angela Chang, Desiree Mitchell, Shreya Minama Reddy, and Hope Schroeder provided invaluable research assistance. J.R. acknowledges the financial support of the Darelyn A. and Richard C. Reed Memorial Fund at the University of Chicago Law School.


